Jahid Hasan

hasan89@purdue.edu | 765-694-5992 | Jahid Hasan | jahid674 | Website | Lafayette, Indiana

Research Interest

Trustworthy Machine Learning, Data Quality, Data Preparation, DBMS, Optimization, Pipeline Robustness, Human-in-the-loop.

Education

Doctor of Philosophy (Ph.D.) — *Computer and Information Technology* Aug 2023 - May 2027
Purdue University, West Lafayette GPA: 3.97
Supervisor: Dr. Romila Pradhan

Master of Science (M.S.) — *Mathematics* Jan 2022 - Jul 2023
University of Dhaka, Dhaka, Bangladesh GPA: 3.81
Dissertation Title: *Age Distributed model of Communicable Disease in Heterogeneous Environment*
Supervisor: Dr. Md. Shahidul Islam

Bachelor Science (B.S.) — *Mathematics* Jan 2017 - Dec 2021
University of Dhaka, Dhaka, Bangladesh GPA: 3.69

Skills

Programming Languages/Frameworks: **Python, MATLAB, SQL, R, Fortran, SAS, SPSS**
Machine Learning Tools/Libraries: **PyTorch, Numpy, Pandas, Tensorflow, Matplotlib, SciKit-Learn**

Work Experience

Purdue University, West Lafayette, IN — *Graduate Research Assistant* Aug 2023 - Present
- Developing a framework to optimize the pipeline for unseen data by utilizing the data profile and historical execution exhibits superior performance over baselines — *Learn2Clean, Bugdoc*.
- Developed a framework that quantifies and explains the robustness of the ML pipeline for the end user. Additionally, the framework automates suggestions for a new pipeline in an efficient manner when the current one fails. [1]
- Developed a Reinforcement Learning-based data expansion framework named *DataSift* for ML training, superior to the baselines – *Random, Entropy, and Autodata* to ensure fairness and accuracy in ML model. Implemented the Influence function in the data pool to assess data value and address uncertainty in utilities. [2]
- Actively working on NSF and CASMI-granted project of *Data Preparation for Fair and Trustworthy Machine Learning*.

University of Dhaka, Bangladesh — *Graduate Research Assistant* Jan 2022 - July 2023
- Derived dynamical behavior and comprehensive data analysis for *'Age Distributed model of Communicable Disease in Heterogeneous Environment'* project granted by the National Science and Technology Department, Bangladesh.

Relevant Project

Purdue University — *Machine Learning Bias Mitigation Through Data Transformation* -
- Implemented a data-transformation method on the raw dataset to ensure equal opportunity in prediction. Demonstrated effective results on the output.

Purdue University — *Heart Attack Prediction: Risk Detection of Cardiovascular Diseases* -
- Developed the entire machine learning pipeline from scratch for the Kaggle Heart Attack dataset.

Involvement

- **CIT Alternate Senator**, Purdue University. Acted as a bridge between my department and the Graduate Government on behalf of graduate students, facilitating communication and representing departmental concerns. $(2024 - 2025)$
- **Event Director**, Polytechnic Graduate Student Organization. Responsible for planning, coordinating, and executing events, including budget management, logistics, and stakeholder engagement. $(2024 - 2025)$

Awards & Professional Activities

- **US Microsoft EDBT Fellowship**, USA, 2025.
- **Participant**, International Conference on Management of Data 2025, **ACM SIGMOD PODS 2025**, Berlin, Germany, 2025.
- **Participant**, EDBT Summer School on AI and Data Management, Nicosia, Cyprus, 2025
- **Travel grant** for the workshop on **Digital Safety**, The Center for Advancing Safety of Machine Intelligence, UL Research Institute. May 2024.
- **National Science and Technology Fellowship (NST)**, Bangladesh, 2022.
  A fellowship by the Science and Technology Ministry of the Government of the People's Republic of Bangladesh for academic excellence.
- **Merit Scholarship**, University of Dhaka, Bangladesh, 2022. A scholarship for 95th percentile academic results.
- **PC Member/Reviewer** in Elsevier 2023 & 2024, ORCiD

Publications

[1] Jahid Hasan and Romila Pradhan. Explanations for machine learning pipelines under data drift. In *Proceedings of the Workshop on Human-In-the-Loop Data Analytics*, HILDA '25, New York, NY, USA, 2025. Association for Computing Machinery.

[2] Jahid Hasan and Romila Pradhan. Data acquisition for improving model fairness using reinforcement learning. *arXiv preprint arXiv:2412.03009*, 2024.

Last Updated: August 18, 2025